



**Syddansk Universitet**

## **SEEDLING DISCRIMINATION USING SHAPE FEATURES DERIVED FROM A DISTANCE TRANSFORM**

Mosgaard Giselsson, Thomas; Jørgensen, Rasmus Nyholm; Midtiby, Henrik Skov

*Publication date:*  
2012

*Document Version*  
Submitted manuscript

[Link to publication](#)

*Citation for published version (APA):*

Mosgaard Giselsson, T., Jørgensen, R. N., & Midtiby, H. (2012). SEEDLING DISCRIMINATION USING SHAPE FEATURES DERIVED FROM A DISTANCE TRANSFORM. Paper presented at International Conference of Agricultural Engineering, Valencia, Spain.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# SEEDLING DISCRIMINATION USING SHAPE FEATURES DERIVED FROM A DISTANCE TRANSFORM

Thomas M. Giselsson<sup>1\*</sup>, Henrik S. Midtiby<sup>1</sup>, Rasmus N. Jørgensen<sup>1</sup>

<sup>1</sup>*University of Southern Denmark, Niels Bohrs Allé 1, Odense, 5000 Denmark*

*\*Corresponding author. Email: [tgi@kbm.sdu.dk](mailto:tgi@kbm.sdu.dk)*

One branch of general image processing research deals with 2D object classification where classes are categorized by different features of the objects such as area, perimeter, elongation, color and texture. When dealing with plant specie classification some of the widely used and well known object features are less useful because the task is to categorize soft objects in outdoor scenes. A general feature set for robust description of soft objects such as plants in an early growth stage is, to our knowledge non existing. We propose a novel way of parametrizing a distance transformation of an object silhouette that may prove to posses value in object classification.

The method approximate the distance distribution of an object with a high degree Legendre polynomial where the polynomial coefficients constitutes a feature set. This feature set will be referred to as Legendre Polynomial Feature Set (LPFS). The method have been tested through a discrimination task where two similar plant species were to be divided into their respective classes. Since the LPFS feature set is meant to be used with a classification algorithm, the performance assessment is the classification accuracy of 4 different classifiers (kNN, Naive-Bayes, Linear SVM, Non-linear SVM). A set of well known features is used for comparison. This feature set will be referred to as Standard Feature Set (SFS). The used dataset consisted of 139 samples of Corn Flower (*Centaurea cyanus* L.) and 63 samples of Night Shade (*Solanum nigrum* L.).

The highest achieved discrimination accuracy with the LPFS feature set was 98.75 % and contained 10 numerical features. The SFS feature set achieved an accuracy of 87.1 % using 22 features. The results show the LPFS feature set can compete with the SFS feature set. Further testing is needed to reveal the true value of the LPFS feature set.

**Key words:** Object discrimination, Machine learning, Feature generation, Precision weeding

## 1. Introduction

Camera based vision systems are one of the technologies that enables the agricultural industry to reduce the amount of chemical use, automate mechanical weeding and assist in gaining detailed information about the weed population and distribution in fields. The technology can rely on systems able to recognize which plant species are present in a field or detect known crop plant patterns (Åstrand & Baerveldt, 2002) (Weis & Gerhards, 2007) (Woebbecke et al., 1995).

One approach is to use dedicated camera systems to capture images in the field and then let computer vision algorithms extract the relevant information present in each image. The image analysis task are non-trivial because it deals with outdoor scenes and the relevant objects are organic and soft in the sense that their appearance varies according to a huge number of factors such as time of day, nutrition and water status, pests in form of insects and fungi, weather conditions, and seasonal variation not to mention the acquisition specific constraints such as viewing angle, blurring, occlusion and light conditions. Further they are continuously changing as they develop.

In this paper we conduct initial studies as to investigate the value of a proposed shape

feature set capable of discriminating between two similar plant species. The final goal is to have a system that can classify common plant species by analysing shapes. This paper documents an attempt to develop shape features minded on plant classification and thereby taking one step toward a general plant specie recognition system.

The paper is structures as follows: section 1 is this introduction to the domain containing also the motivation for developing systems with plant discrimination capabilities. Section 2 documents the used data sets and describe the feature generation results. Section 3 states the achieved results by applying classifiers to the generated feature sets. Section 4 analyses the classification errors made by the best performing classifier and suggest pros and cons of the proposed LPFS feature set by consulting the observed errors. Appendix A and B contains the full data set used in this study.

## 2. Materials and methods

### 2.1 Data

The data used for this paper consist of 139 image samples of Corn Flower (*Centaurea cyanus*) and 63 image samples of Night Shade (*Solanum nigrum*). Both plant species vary in growth stages between BBCH10 and BBCH11<sup>1</sup>.

The plant specie and growth stage combination is chosen for two reasons. (a) The data was available in a preexisting data base and (b) these specific species are very alike from a human point of view while in dicotyledon stage. The common growth stage can be described as the time span from when the two cotyledons are fully developed and one or two leaves are emerging.

Plant samples have been extracted from images of potted plants. Each pot contained a single plant which helped automatic segmentation and annotation. The entire object extraction procedure is described thoroughly in Giselsson (2010). To give the reader the best possible insight into the data set the authors have chosen to include every sample in appendix A and B.

### 2.2 Feature generation

Binary images were constructed from the result of the mentioned object extraction process. With data in the form of binary connected components several features could be calculated. The calculated features are split up into two groups. The first set consisted of 21 numeric features widely used in the literature (e.g. in Du et al., (2007), Åstrand & Baerveldt (2002), Woebbecke *et al.*, (1995), Mei (2010), Weis & Gerhards (2007)) of object recognition: Seven invariant image moments; Object area; Object perimeter; Convex hull area; Solidity; Convex hull perimeter; Perimeter ratio; Eccentricity; Compactness; Circular variance; Elliptic variance; and four skeleton based features. This set is referred to as the Standard Feature Set (SFS). The second set consisted of the proposed novel shape features. This set is referred to as the Legendre Polynomial Feature Set (LPFS) and is in essence a parametrization of a distance map generated from binary object images.

Having a binary image of an object the features are generated by first constructing the distance map. All distances corresponding to object pixels are collected into a list that is then sorted. Finally the list of distances are scaled such that the largest distance equals 1 which makes the calculated features robust against scale changes. Now a 10 degree Legendre polynomial is fitted to the numbers in the list and the computed coefficients constitute the LPFS features. An elaborated description of the LPFS features can be found in Giselsson (2012).

---

<sup>1</sup> For a description of growth stage notation see (JKI, O. 2001)

Well performing generic shape features used in computer vision often possess three properties of invariance. These are their ability to be invariant to scale, rotation and translation. Having features with such properties and a robust object extraction method, objects can be recognized as long as they have not undergone deformation. Another way of saying this is to demand that objects are only subject to similarity transformations. Some features claim also to be invariant to small affine transformations (Mei, 2010) (Dionisio & Kim, 2004).

The case studied in this paper employ standard shape features that are rotation, translation and scale invariant. The objects considered in this case study are not static so requiring features to have the mentioned invariant properties are only a minimal requirements and do not suffice. An ideal feature in the context of this study would be invariant to in-class natural variations. The existence of such features are unknown to the author. The proposed feature set introduces one additional invariance with respect to object deformation that is of specific importance when the task is to recognize plants. Consider a seedling with 2 cotyledons as seen in figure 1.a.

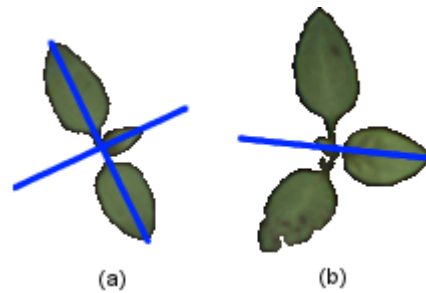


FIGURE 1: (a) Example of seedling with 2 symmetry axes. (b) Example of seedling with one symmetry axis

These two leaves emerge pairwise in the studies species. At this point the seedling has 2 symmetry axes. The next stage starts with the emergence of the first true leaf. This can be seen in figure 1.b. Now the plant only has one symmetry axis which can be described by saying that the relative stem angles have changed. The fact that stem angles change according to growth stage or is different depending on the sample in question is an example of in-class natural variation. The proposed LPFS features do not change because of this angle change as long as the angle change do not lead to overlap. This property is inherited from the distance transform that forms the basis of the data preprocessing.

### 2.3 Feature set quality assessment

The quality of the proposed feature sets are assessed by the performance of classifiers using LPFS and compared with the results of a classifier using the SFS feature set.

To avoid numerical problems the feature values were translated and scaled to be contained in the interval  $[-1;1]$  before use. Four classifiers was tested: 3-Nearest Neighbor, Naive Bayes, Linear SVM and a RBF-SVM classifier. Classifier parameter tuning and classification was conducted by the matlab toolbox PRtools (Pekalska et al., 2007). The performance of the classifier is measured by its accuracy:

$$Accuracy = (TP + TN) / N$$

Where  $TP$  is true positives,  $TN$  is true negatives and  $N$  is total number of samples.

The used classifiers were all supervised learning machines and needed training. The division of the samples into a training and a test set was accomplish by stratified cross validation to ensure that no samples are used both for training and testing and that the class distribution in the two sets are alike. The cross validation was performed using 10 folds.

### 3 Results

A total of eight classification results were collected. This number emerge from having four classifiers and two datasets. The results can be seen in table 1. Using the RBF SVM classifier and letting PRTools perform parameter tuning an accuracy of 97.5 was achieved.

TABLE 1 Rounded accuracies of the four applied classifiers on each feature set

	3-Nearest Neighbor	Naive Bayes	Linear SVM	RBF SVM
SFS	89 %	74 %	93 %	86 %
LPFS	95 %	90 %	92 %	98 %

### 4 Discussion of classification errors

The author believes that it gives important insight to investigate which kind of errors are made by the applied classifiers. Because of the abstract nature of the proposed distance features and the high dimensionality of the space they occupy the analysis of classification errors will be conducted by visually inspecting errors and comparing those with the true class samples.

Figure 2 shows the five samples that was erroneously classified by a RBF-SVM classifier. Comparing those to other class samples in appendix A and B reveals that the cornflower samples in figure 2.a seems to either be very early in their development stage or being in an (for this data set) unusually pose. The nightshade sample in figure 2.b is in a stage where the two cotyledons have developed and a true leaf is on its way but in this particular situation the true leaf overlap one of the cotyledons changing the object outline dramatically compared to the general shape of the rest of the samples.

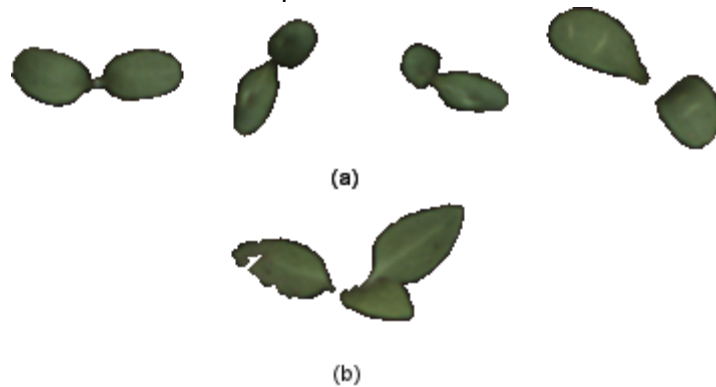


FIGURE 2: Samples that were not correctly classified. (a) Erroneous classified cornflower samples. (b) Erroneous classified nightshade samples

Acknowledging the errors raise two questions: Is it realistic to assume that a feature set based on the shape of whole plant canopies can be made to be invariant to natural variation of species even if a limited range in growth stages are required? Is the classification accuracy reported in this paper acceptable?

It is the authors belief that shape features of whole canopies can not be made general in such a degree that all natural variations are covered. Having a large sample database might help, but could also result in excessive noise. The question if the classification accuracy will suffice will depend on the particular case. For weed maps or the task of making decisions on herbicide mixtures and dosis not every plant needs to be recognized and therefore the ability of the presented system will suffice.

## Conclusion

Using a novel set of shape describing features called LPFS it was possible to discriminate between seedlings of Corn flower (*Centaurea cyanus*) and Night Shade (*Solanum nigrum*). By using an RBF SVM classifier an accuracy of nearly 98 % was achieved. If the approach can be applied to additional plant species with similar positive results the system can be used in a decision making process to determine appropriate herbicide application and construct weed maps.

## Reference list

- Åstrand, B., & Baerveldt, A.-jan. (2005). Plant recognition and localization using context information and individual plant features. Mechatronics and Robotics 2004 special session Autonomous Machines in Agriculture.
- Weisa, M., Gerhards, R., Stafford, J., & others. (2007). Feature extraction for the identification of weed species in digital images for the purpose of site-specific weed control. Precision agriculture, 2007, 537–544.
- Woebbecke, D. M., Meyer, G. E., Von Bargen, K., & Mortensen, D. A. (1995). Shape features for identifying young weeds using image analysis. Transactions American Society of Agricultural Engineers ASAE, 38(1), 271–281.
- Giselsson, T. M. (2010). Real Time Crops and Weeds Classification from Top Down Images of Plant Canopies. Master Thesis, University of Southern Denmark
- JKI, O. (2001). Growth stages of mono-and dicotyledonous plants. BBCH Skala.
- Giselsson, T. M. (2012). Seedling Discrimination using Shape Features derived from a Distance Transform. (Submitted to Sensors).
- Mei, Y. (2010). Robust affine invariant shape descriptors. Ph.D Thesis. Ryerson University (Canada).
- Dionisio, C. R. P., & Kim, H. Y. (2004). New features for affine-invariant shape classification. Image Processing, 2004, 2135–2138
- Pekalska, E., Ridder, D. D., Tax, D. M. J., & Verzakov, S. (2007). A Matlab Toolbox for Pattern Recognition. Pattern Recognition, 1(August), 1-61

**Appendix A** - 63 image samples of Night Shade (*Solanum nigrum*). The red circle is the misclassified samples using the LPFS feature set and RBF-SVM classifier

